

EXTENSION PLAN - DATA 512

PRERIT CHAUDHARY

11/17/2023

MOTIVATION / PROBLEM STATEMENT

Smoke fires have a huge impact on a number of aspects of a city. It can be seen that the number of fires have increased over the last few years and have a significant impact on the environment. Urban and rural landscapes are impacted in many different ways. Urban landscapes suffer a more significant impact due to the density of people living in the city which affects both the quality of human lives and overall infrastructure. Beyond the impact of fires on an immediate level, they also have long term consequences on healthcare, environment and economic aspects of a city. People living in the city may see an increase in respiratory issues. Apart from its effect on health, fires also have repercussions on the economic conditions through property damage, disruption of businesses, agricultural losses and tourism. The city council needs to have a proper plan to address all these concerns and make people aware of the measures that they have taken to overcome these losses. The aim of my analysis would be around understanding the link and correlation between fires and mortality rates due to respiratory diseases over the last few years. Smoke fires can introduce different particulate matter and gases into the air which can cause a number of respiratory issues which can have short term and long term consequences. This can impact not just the residents with existing respiratory conditions but can also impact other people not with pre-existing conditions as well. The analysis would be done on an overall level where I would be monitoring the change in mortality rate due to respiratory diseases. Subsequently, a root cause analysis will be conducted to identify the target resident groups most susceptible to respiratory diseases. This analysis will involve analyzing different age groups, gender, and ethnicity to identify trends and patterns.

This analysis is important to understand the long term consequences of fires on health and finding correlations and building a predictive model would help generate meaningful insights from this. These insights can be used to create action plans and recommendations for the city council and mayors regarding policy changes that need to be made to safeguard the highly prone residents. The true essence of the analysis lies in the human centered approach while conducting this analysis. This might involve focusing on finding correlations and statistical significance of hypotheses, but the narrative focuses on improving the well being of the people living in Derby, Kansas. The end goal of the analysis would be to generate actionable insights that could bring about

a change in the policies to safeguard people living in Derby and create a tangible impact to the quality of their lives.

IMPACT FOCUS

The primary focus of this analysis is to understand the impact of smoke fires on the health of residents living in Derby, Kansas. The area of focus is Health care as it is important to identify key trends and patterns so as to bring about a change for the betterment of the residents living there. The analysis would involve combining the previously built predictive model for smoke fires in the near future with the new data sources identified to predict the mortality rates due to respiratory diseases in the coming years. The analysis would revolve around answering the following questions around the impact of smoke fires on residents in Derby, Kansas -

1. Mortality Rates and Respiratory Diseases -

- What is the correlation between increase in smoke fires and mortality rates due to respiratory conditions
- Which respiratory diseases are shown to have a significant increase in mortality rates and if they have the highest correlation with the increase in smoke fires

2. Demographic Disparity

- Identifying age groups most vulnerable to respiratory health conditions
- Identifying cohorts in sex and ethnicity which are most prone to health issues due to fire

DATA USED

Data Source -

- <https://ghdx.healthdata.org/record/ihme-data/united-states-chronic-respiratory-disease-mortality-rates-county-1980-2014>
- <https://ghdx.healthdata.org/record/ihme-data/united-states-causes-death-life-expectancy-by-county-race-ethnicity-2000-2019>

The above two data sources include most of the information that would be needed to address the research questions and validate the statistical significance of our results

License and Terms of Use

The data is available on the GHDx website which is the world's most comprehensive catalog of surveys, censuses, vital statistics, and other health-related data. The data is open source and widely accessible to the users and contains all the information available for any healthcare analysis

Data Profiling

Data Source 1

Mortality rates aggregated at a five year interval for different counties in Kansas with their FIPS between the time period 1980 - 2014.

FIPS - The Federal Information Processing Standards (FIPS) code, a unique identifier for states and counties in the United States

Data is available for the following respiratory conditions -

- Chronic respiratory diseases
- Chronic obstructive pulmonary
- Pneumoconiosis
- Silicosis
- Asbestosis
- Coal workers pneumoconiosis
- Other pneumoconiosis
- Asthma
- Interstitial lung disease
- Other chronic respiratory

This data source will help address the first question which focuses on finding a correlation between smoke fires and overall mortality rates due to respiratory conditions. The analysis would further review if there are certain respiratory diseases which have seen a major increase as compared to others and how their correlation is with increase in smoke fires.

Data Source 2

This data contains a detailed view of the life expectancy and mortality rates due to respiratory conditions with columns such as Location, FIPS, Race, Sex, Age group due to respiratory infections and chronic respiratory conditions. The estimates are given in the following three columns :

- Val - Estimate for that specific county
- Lower - 2.5% percentile estimate
- Upper - 97.5% percentile estimate

This data source will help find patterns across different demographics to identify highly prone risk groups who are being affected the most. The analysis would initially be done at a univariate level for each demographic data such as race, sex and age. This would be followed by a bivariate analysis to check for combination of these three attributes.

UNKNOWNNS AND DEPENDENCIES

There are various unknown factors that might impact the credibility and reliability of the results of the analysis. We are only considering two variables from a subset of variables which might be correlated. It is also a well known fact that there is a significant difference between correlation and causation and hence any results from this analysis would be just a high level view of the relationship between smoke fires and respiratory conditions. There might be other variables such as individuals with preexisting conditions, socioeconomic factors and access to healthcare that can be confounding variables in our analysis. The data under consideration is specific to a county and hence any trends might be at a local level and cannot be extrapolated as is to a global population. Also, data quality is unknown as there is limited information available on the method of data collection and accuracy of data reported.

TIMELINE TO COMPLETION

- **November 17, 2023 - Submission of Extension Plan**
- **November 20, 2023 - Exploratory Data Analysis**
- **November 24, 2023 - Correlation Analysis and Feature Engineering**
- **November 27, 2023 - Model Building**
- **November 29, 2023 - Compiling Results and Validating Insights**
- **November 30, 2023 - Submission of Presentation Deck**
- **December 11, 2023 - Repository and Report Submission**

REFERENCES

1. D'Evelyn SM, Jung J, Alvarado E, Baumgartner J, Caligiuri P, Hagmann RK, Henderson SB, Hessburg PF, Hopkins S, Kasner EJ, Krawchuk MA, Krenz JE, Lydersen JM, Marlier ME, Masuda YJ, Metlen K, Mittelstaedt G, Prichard SJ, Schollaert CL, Smith EB, Stevens JT, Tessum CW, Reeb-Whitaker C, Wilkins JL, Wolff NH, Wood LM, Haugo RD, Spector JT. Wildfire, Smoke Exposure, Human Health, and Environmental Justice Need to be Integrated into Forest Restoration and Management. *Curr Environ Health Rep.* 2022 Sep;9(3):366-385. doi: 10.1007/s40572-022-00355-7. Epub 2022 May 7. PMID: 35524066; PMCID: PMC9076366. (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9076366/>)
2. <https://www.epa.gov/wildfire-smoke-course/which-populations-experience-greater-risks-adverse-health-effects-resulting>